



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

RPG-SAM: Reliability-Weighted Prototypes and Geometric Adaptive Threshold Selection for Training-Free One-Shot Polyp Segmentation

Weikun Lin^{1*}, Yunhao Bai^{1*}, and Yan Wang^{1†}

Shanghai Key Laboratory of Multidimensional Information Processing,
East China Normal University

10232140464@stu.ecnu.edu.cn, ywang@cee.ecnu.edu.cn

*Equal contribution. †Corresponding author.

Abstract. Training-free one-shot segmentation offers a scalable alternative to expert annotations by transferring knowledge from a labeled support image to query images with the help of foundation models. However, existing methods often treat support regions and query response intensities homogeneously, ignoring regional heterogeneity in the support image and response heterogeneity across query images. To address these limitations, we propose RPG-SAM, a training-free framework for one-shot polyp segmentation. Specifically, Reliability-Weighted Prototype Mining (RWPM) prioritizes high-fidelity support features and uses background anchors as contrastive references for noise suppression. Geometric Adaptive Selection (GAS) dynamically recalibrates binarization thresholds by evaluating the morphological consensus of candidate masks. Finally, Prior-guided Iterative Refinement (PIR) uses SAM2 to refine anatomical boundaries. By explicitly modeling multi-level information heterogeneity, RPG-SAM achieves a 5.56% mIoU improvement on the Kvasir dataset. Code is publicly available at <https://github.com/AlwaysHF/RPG-SAM>.

Keywords: Polyp Segmentation · Colonoscopy · Segment Anything Model

1 Introduction

Polyp detection is a key task in early colorectal cancer (CRC) screening[20, 17, 7]. Although supervised models have achieved high accuracy[5, 6, 15], their dependence on large-scale pixel-level annotations limits clinical scalability. Recently, training-free one-shot segmentation[12, 19, 3, 10, 21] has emerged as a promising alternative, integrated with vision foundation models[13, 14], particularly the Segment Anything Model (SAM) [9, 16]. SAM’s proven zero-shot generalization and promptable interface allow for high-quality mask generation from a single labeled *support* image to *query* images without any parameter updates, offering a fast solution for label-scarce medical scenarios.

The standard workflow for training-free one-shot segmentation methods follows a matching-based prompt sampling paradigm. This involves two core stages:

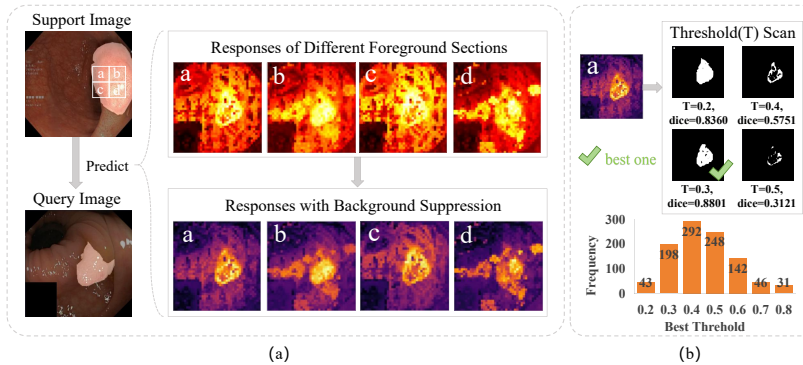


Fig. 1. Motivation of RPG-SAM. (a) Regional and Contextual Heterogeneity. Top: The varying regional utility of the support foreground means that degraded areas (e.g., reflection *d*) trigger false-positive noise in query images. Bottom: Treating the contextual background as a distinct information layer provides negative anchors to effectively suppress heatmap activations. (b) Intensity Heterogeneity. The histogram of optimal binarization thresholds across the Kvasir dataset [8] highlights the stochastic response intensities of different query scenarios, exposing the limitations of fixed-threshold (homogeneous) sampling rules.

1) a feature matching stage that computes a similarity heatmap between the query image and support foreground features, and 2) a prompt sampling stage that applies various strategies to extract points, boxes, or masks for SAM input. For instance, PerSAM[21] and Matcher[10] sample positive and negative points based on similarity extremes or bidirectional consistency; while OPSAM[11] and ProtoSAM[19] binarize heatmaps using fixed thresholds or regional statistics to extract candidate masks. These manual rules serve as the bridge between raw feature matching and the foundation model’s prompt interface.

However, we observe that the effectiveness of this paradigm is fundamentally limited by a uniformity bias, where current pipelines erroneously assume homogeneity across three critical dimensions of information. First, regarding regional homogeneity within support foreground, existing methods often treat all support foreground pixels as equally representative. In practice, colonoscopic images frequently contain suboptimal regions, such as those obscured by reflections or mucus, that introduce misleading features if not explicitly filtered. As illustrated in the top of Fig. 1(a), matching features from different regions yields heatmaps of varying quality; specifically, degraded regions (e.g., region *d*) trigger extensive false-positive noise. Second, regarding regional homogeneity across foreground and background, many methods neglect the support background as a distinct information layer, failing to leverage it as an explicit contrastive reference. As shown in the bottom of Fig. 1(a), utilizing background features as negative anchors is essential for suppressing noise and distinguishing polyps from similar-looking intestinal folds. Finally, regarding intensity heterogeneity, the transition from heatmaps to prompts is currently governed by static rules that ignore the

stochastic response intensities across diverse query images. As evidenced by the histograms in Fig. 1(b), the optimal binarization thresholds fluctuate significantly, making rigid sampling rules inadequate for ensuring both prompt fidelity and diversity in varied clinical conditions.

To mitigate this uniformity bias and ensure high-fidelity knowledge transfer, we propose RPG-SAM, a robust framework that explicitly accounts for information heterogeneity. We first introduce Reliability-Weighted Prototype Mining (RWPM) to resolve regional heterogeneity within the support foreground and across foreground and background; it distills support features according to their intrinsic reliability while leveraging background anchors as explicit contrastive references. To address intensity heterogeneity, we develop Geometric Adaptive Selection (GAS), which replaces static sampling rules with a dynamic thresholding mechanism guided by morphological priors. Finally, a Prior-guided Iterative Refinement (PIR) loop is designed to progressively polish anatomical boundaries. By systematically tackling these heterogeneous cues, RPG-SAM maximizes the utility of sparse support information and maintains high robustness across diverse endoscopic conditions. RPG-SAM focuses on the training-free one-shot setting, where no dataset-specific optimization or task-specific embedder training is performed, making it complementary to recent exemplar-prompted foundation-model extensions.

2 Methodology

RPG-SAM is a SAM2-based training-free framework designed to segment polyps in a query image x_q using a single labeled support image x_s and its corresponding mask m_s . As illustrated in Fig. 2, the framework consists of three main components: Reliability-Weighted Prototype Mining (RWPM), Geometric Adaptive Selection (GAS), and Prior-guided Iterative Refinement (PIR). We introduce the components in the following subsections.

Let x_s and m_s denote the support image and its binary mask, and x_q denote the query image. DINOv2 extracts support and query feature maps $f_s, f_q \in \mathbb{R}^{H'W' \times D}$. Foreground and background prototypes are mined from f_s according to m_s , and the resulting query response map is denoted as $\mathbf{H}_{init}$.

2.1 Reliability-Weighted Prototype Mining

To address regional and contextual heterogeneity, we utilize DINOv2[13] as the feature extractor. The support image x_s is divided into foreground and background regions according to m_s . To handle the complexity of endoscopic scenes, we apply the SLIC[1] algorithm to the support image to generate K super-pixel clusters. By averaging features within these clusters according to m_s , we obtain the foreground prototypes $\mathbf{P}_{fg} = \{p_{fg}^k\}_{k=1}^K$ and background prototypes $\mathbf{P}_{bg} = \{p_{bg}^k\}_{k=1}^K$.

To distinguish the reliability of $\mathbf{P}_{fg}$, we introduce a module that evaluates each $\mathbf{P}_{fg}^k$ based on two complementary metrics: **Intrinsic Reliability** (Contrast

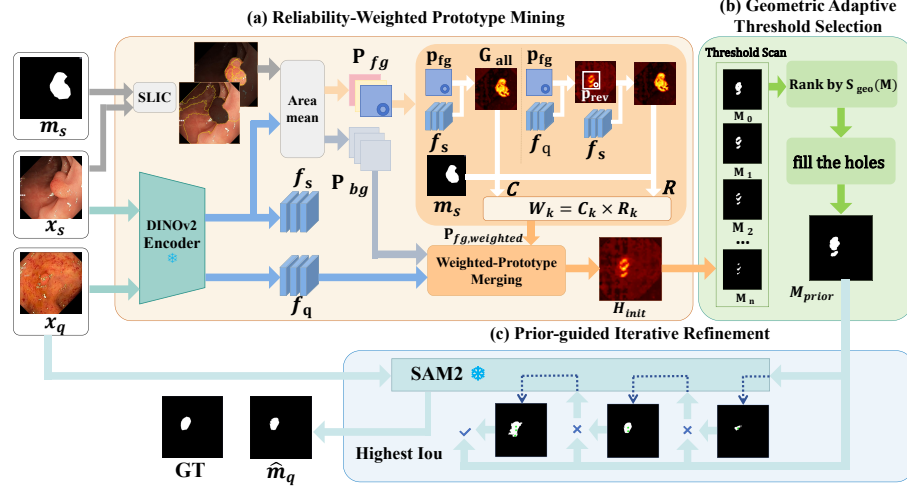


Fig. 2. Overview of RPG-SAM. Given a support image-mask pair and a query image, RPG-SAM first extracts DINOv2 features and mines reliability-weighted foreground prototypes with background suppression. It then performs geometric adaptive selection to obtain a query-specific prior mask, which is finally refined by SAM2 through prior-guided iterative prompting.

Factor) and **Query-specific Relevance** (Reverse Purity Factor). **First**, we compute the Contrast Factor $\mathbf{C}_k$ to evaluate discriminative power of a prototype within the support image. We compute the similarity between p_{fg}^k and all support features f_s , put the similarity scores $\mathbf{G}_{all}$ into foreground sets $\mathbf{G}_{fg}$ or background sets $\mathbf{G}_{bg}$ according to the mask m_s . The contrast is defined as the standardized difference between these sets:

$$\mathbf{C}_k = \text{ReLU} \left(\frac{\text{mean}(\mathbf{G}_{fg}) - \text{mean}(\mathbf{G}_{bg})}{\text{std}(\mathbf{G}_{all})} \right). \quad (1)$$

Prototypes that fail to distinguish themselves from the support background (e.g., those corrupted by specular reflections) receive a lower $\mathbf{C}_k$. **Second**, we use a Reverse Purity Factor $\mathbf{R}_k$ to verify the cross-image matching stability of $\mathbf{P}_{fg}^k$. We identify its top- n matches in the query feature map f_q , average them to form a proxy p_{rev}^k , and project this proxy back onto f_s . The purity p is the proportion of top- n reverse matches falling within m_s . To account for varying lesion scales, we normalize this purity against the foreground area ratio p_0 :

$$\mathbf{R}_k = \text{ReLU} \left(\frac{p - p_0}{1 - p_0} \right). \quad (2)$$

The ReLU operation serves as a reliability gate. A negative contrast score indicates that the foreground prototype is not more discriminative than the support background, while a negative reverse-purity score indicates that its reverse

matches are no better than the foreground area prior. Such unreliable prototypes are therefore suppressed instead of being assigned unstable negative weights. We dynamically weight each prototype’s contribution based on its relevance to the current query, where $\mathbf{W}_k = \mathbf{C}_k \cdot \mathbf{R}_k$.

Finally, we generate an initial heatmap $\mathbf{H}_{init}$ by aggregating the weighted foreground similarity maps. Unlike methods that focus solely on foreground cues, we explicitly leverage $\mathbf{P}_{bg}$ as negative anchors to suppress false-positive activations in the heatmap:

$$\mathbf{H}_{init} = \text{Softmax} \left(K_{fg} \sum_{k=1}^{K_{fg}} W_k \cdot (f_q \cdot p_{fg}^k)^\top - \sum_{k=1}^{K_{bg}} (f_q \cdot p_{bg}^k)^\top \right). \quad (3)$$

To enhance spatial consistency and reduce isolated noise, we perform affinity-based iterative heatmap refinement[11]. Specifically, we construct a query-query affinity matrix from DINOv2 features and propagate the initial heatmap over this graph, so that feature-similar pixels reinforce each other while isolated activations are suppressed:

$$\mathbf{A}_q = \text{Softmax}(f_q f_q^\top), \quad \mathbf{H}^{(t+1)} = \mathbf{A}_q \mathbf{H}^{(t)}, \quad t = 0, \dots, N-1, \quad (4)$$

where $\mathbf{H}^{(0)} = \mathbf{H}_{init}$, and the final refined heatmap is used for subsequent geometric adaptive selection.

2.2 Geometric Adaptive Selection

To mitigate the intrinsic stochasticity of response intensities across different query images, we develop Geometric Adaptive Selection (GAS), which performs query-adaptive threshold selection. Instead of using a fixed binarization threshold, GAS dynamically evaluates a pool of candidate masks to identify the one that best conforms to the expected geometric properties of a polyp.

We first generate a set of candidate binary masks $\{\mathbf{M}_\tau\}$ by thresholding the initial heatmap $\mathbf{H}_{init}$ across a range of potential levels $\tau \in [\tau_{min}, \tau_{max}]$. For each $\mathbf{M}_\tau$, we perform connected component analysis and retain only those components whose area is at least 20% of the largest component in that mask. To ensure morphological completeness, internal holes within these valid regions are filled, resulting in refined candidate masks. Then, each refined candidate $\mathbf{M}$ is evaluated by a geometric score $\mathbf{S}_{geo}$, which balances morphological solidity and scale consensus:

$$\mathbf{S}_{geo}(\mathbf{M}) = \underbrace{\sum_i \frac{|\mathbf{C}_i|}{|\mathbf{M}|} \frac{|\mathbf{C}_i|}{|\text{Hull}(\mathbf{C}_i)|}}_{\text{Weighted Solidity}} \cdot \underbrace{\min \left(1, \frac{|\mathbf{M}|}{\mathbf{A}_{ref}} \right)}_{\text{Scale Consensus}},$$

where $\mathbf{C}_i$ denotes the i -th valid component of $\mathbf{M}$, and $|\cdot|$ represents the area. The weighted solidity computes the area-weighted average compactness of the

components, favoring regions with regular, convex-like anatomical shapes. To prevent small, fragmented noise from obtaining artificially high solidity scores, we introduce Scale Consensus. This term downweights candidates that are significantly smaller than a reference area $\mathbf{A}_{ref}$, which represents the expected median scale of polyps. The candidate mask with the highest $\mathbf{S}_{geo}$ is selected as the optimal prior mask $\mathbf{M}_{prior}$ to generate sparse prompts for SAM2.

2.3 Prior-guided Iterative SAM2 Refinement

To fully automate the segmentation process without manual intervention, we implement an iterative refinement strategy that utilizes the geometric prior $\mathbf{M}_{prior}$ to guide SAM2. This process progressively optimizes the segmentation mask $\mathbf{M}_t$ by using the structural consistency of the prior as a reference for error correction. In each iteration, the current mask is evaluated against $\mathbf{M}_{prior}$ based on Coverage (Cov) and Intersection-over-Union (IoU), allowing the framework to utilize SAM2’s edge-refinement capabilities to polish anatomical boundaries.

The refinement logic follows a hierarchical approach by first addressing insufficient lesion coverage ($Cov < \tau_{cov}$). When such errors are detected, the system identifies the false-negative region $\mathbf{R}_{FN} = \mathbf{M}_{prior} \cap \neg \mathbf{M}_t$ and samples its geometric center via the Euclidean Distance Transform (EDT) as a positive prompt to expand the mask. Conversely, if coverage is adequate but the IoU remains sub-optimal, the focus shifts to suppressing background noise by identifying the false-positive region $\mathbf{R}_{FP} = \mathbf{M}_t \cap \neg \mathbf{M}_{prior}$ and inserting a negative prompt. This iterative correction continues until the stopping criteria $(Cov_t \geq \tau_{cov}) \wedge (IoU_t \geq \tau_{iou})$ are satisfied or the maximum iteration count T_{max} is reached. To ensure robustness, the mask with the highest IoU relative to the prior throughout the iteration history is selected as the final query prediction $\hat{m}_q$.

3 Experiments and Evaluation

3.1 Experiment Settings

Datasets and Metrics. We validate the effectiveness of RPG-SAM on four widely used public polyp segmentation datasets: **Kvasir**[8], **CVC-ClinicDB**[4], **CVC-ColonDB**[18] and three-center **PolypGen**[2]. Following OPSAM[11], we randomly select one image as the support sample and keep it fixed for all methods to ensure fair comparison. We employ the mean Dice coefficient (mDice) and mean Intersection over Union (mIoU) as the primary evaluation metrics.

Implementation Details. Our framework is implemented in PyTorch. We use DINOv2[13] with a ViT-L/14 as the feature encoder with an input resolution of 560×560 , and SAM2[16] as the mask generator with an input size of 1024×1024 . The hyperparameter settings are as follows: (1) For **RWPM**, the SLIC parameters are set to $K = 10$ and $m = 20$. (2) For **GAS**, the confidence scanning range is set to $[0.4, 0.7]$ with a stride of 0.05. (3) For the **PIR**, the stopping thresholds are set to $\tau_{cov} = 0.9$ and $\tau_{iou} = 0.8$, with a maximum iteration count of $T_{max} = 5$. All inference processes are conducted on a single NVIDIA RTX 3090 GPU without any model training or fine-tuning.

Table 1. Comparison with state-of-the-art training-free one-shot segmentation methods across public polyp segmentation datasets. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	Kvasir		PolypGen						CVC-ClinicDB		CVC-ColonDB	
	IoU	Dice	Center 1		Center 2		Center 3		IoU	Dice	IoU	Dice
SEGIC[12]	22.23	32.81	13.36	20.73	15.47	22.79	15.73	23.77	17.87	26.98	12.00	18.25
PerSAM[21]	26.01	35.38	26.15	31.47	12.97	18.96	10.87	17.74	13.29	20.10	10.31	16.18
OPSAM[11]	63.95	71.43	<u>57.16</u>	62.53	59.29	65.33	<u>74.91</u>	<u>81.67</u>	58.75	65.60	<u>56.44</u>	62.86
Matcher[10]	71.71	78.91	56.25	<u>63.06</u>	<u>60.79</u>	<u>66.30</u>	65.49	71.42	63.39	70.20	52.01	58.42
ProtoSAM[3]	<u>73.09</u>	<u>81.54</u>	51.14	58.63	55.73	63.03	69.79	76.96	<u>68.99</u>	<u>76.83</u>	55.92	<u>63.45</u>
RPG-SAM (Ours)	78.65	85.65	59.45	64.78	61.04	67.14	76.75	83.56	70.18	77.76	61.16	68.77

3.2 Comparison with Baseline Models

We compare RPG-SAM against SOTA one-shot methods, including **SEGIC**[12], **PerSAM**[21], **Matcher**[10], **OPSAM**[11], and **ProtoSAM**[3]. For fairness, all methods utilize identical support-query pairs and backbone settings. Results are detailed in Table 1. RPG-SAM consistently outperforms existing paradigms. Notably, on Kvasir, it achieves 78.65% mIoU and 85.65% mDice, surpassing ProtoSAM by 5.56% and 4.11%, respectively. This improvement stems from RWPM, which discriminatively prioritizes reliable support features and suppresses misleading features caused by specular reflections, unlike methods assuming uniform representation. On the multi-center PolypGen dataset, RPG-SAM demonstrates superior robustness. While domain shifts often trigger false-positive activations in other models, our Background Suppression and GAS stabilize response intensities across centers. This suggests that RPG-SAM is more robust under cross-center domain shifts. Unlike fixed-threshold methods (e.g., OPSAM), the GAS module adaptively recalibrates for each query image. This prevents prompt contamination in low-contrast cases, ensuring the iterative refinement loop operates on high-quality, diverse candidates to polish anatomical boundaries.

3.3 Ablation Studies

We evaluate the components and hyperparameter sensitivity of RPG-SAM on the Kvasir dataset. As shown in Table 2, each module contributes significantly to the final performance. The baseline without background suppression performs poorly due to unconstrained feature transfer. Explicitly introducing Background Suppression (BG Supp.) yields a 3.78% gain in mDice, proving its necessity in penalizing false-positive activations. RWPM further improves spatial granularity, relieving the uniform representation fallacy. The GAS module outperforms fixed thresholds ($\tau = 0.7$, optimal) by 2.59% mDice, effectively mitigating response intensity stochasticity. Finally, PIR further improves the results by refining boundary details with SAM2. AUC-PR is computed on the heatmap before mask selection and SAM2 refinement; therefore, it remains unchanged for variants that only modify downstream selection or refinement stages.

We further analyze the robustness of hyperparameters in Table 3. For GAS, the scanning range $[0.4, 0.7]$ provides the optimal candidate pool; lower ranges

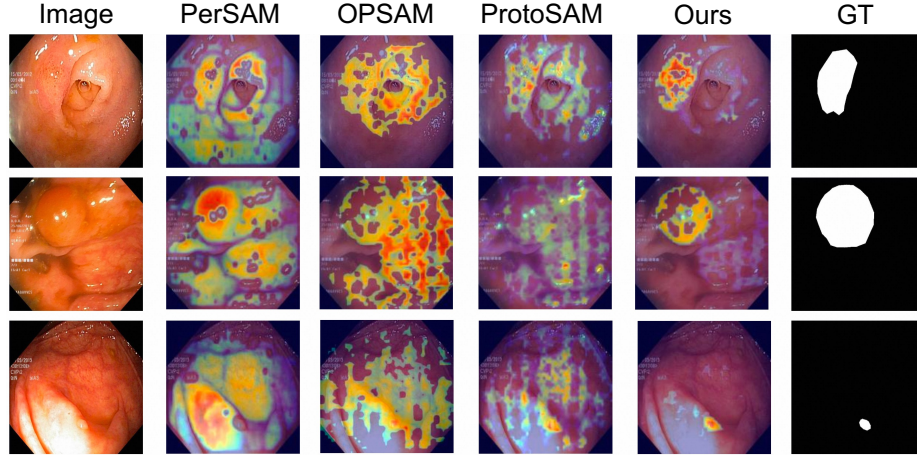


Fig. 3. Visualization of heatmaps on challenging query samples, including cases with small ground-truth regions or the presence of visually similar distractors.

Table 2. Incremental ablation on Kvasir dataset. We demonstrate RPG-SAM benefits of Background Suppression (BG Supp.), RWPM, GAS, and PIR.

BG Supp.	RWPM	GAS	PIR	mIoU (%)	mDice (%)	AUC-PR
-	-	-	-	63.68	74.94	0.8743
✓	-	-	-	67.89	78.72	0.8921
✓	✓	-	-	71.27	81.05	0.9068
✓	✓	✓	-	75.25	83.64	0.9068
✓	✓	✓	✓	78.65	85.65	0.9068

introduce excessive noise, while higher ranges limit diversity. For RWPM, SLIC parameters $K = 10$ and $m = 20$ strike the best balance between spatial accuracy and regularization. Finally, the framework exhibits high stability regarding PIR thresholds, with $\tau_{cov} = 0.9$ and $\tau_{iou} = 0.8$ achieving the optimal balance between recall and precision.

4 Conclusion

We proposed RPG-SAM, a training-free framework addressing noise and intensity stochasticity in one-shot polyp segmentation. RWPM resolves the uniform representation fallacy through reliability-driven feature distillation and background suppression, while GAS replaces rigid thresholds with adaptive binarization based on geometric priors. Extensive benchmarks, including multi-center cohorts, demonstrate the effectiveness of RPG-SAM in terms of generalization and enhanced performance. Our approach provides a robust, scalable alternative to data-intensive models in label-scarce clinical settings. Future work will fur-

Table 3. Ablation study on hyper-parameters for SLIC, GAS, and PIR modules. Params in **bold** are used in RPG-SAM.

SLIC params				PIR params				GAS params		
K	m	mIoU	mDice	τ_{cov}	τ_{iou}	mIoU	mDice	Range	mIoU	mDice
5	5	77.34	84.52	0.8	0.7	77.98	85.20	0.3-0.6	75.05	82.57
10	5	78.28	85.27	0.8	0.8	78.11	85.32	0.4-0.6	77.15	84.48
10	20	78.65	85.65	0.8	0.9	77.79	85.12	0.4-0.7	78.65	85.65
10	50	78.34	85.41	0.9	0.8	78.65	85.65	0.5-0.7	77.26	84.57
20	5	76.08	83.21	0.95	0.8	78.66	85.60	0.5-0.8	77.12	84.41

ther investigate its generalization to broader anatomical structures and imaging modalities, stratified performance under clinical factors such as image quality, bowel preparation, polyp morphology, size, and type, and temporal consistency in endoscopic videos.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 62471182), the Science and Technology Commission of Shanghai Municipality Basic Research Program (Grant No. 25JD1401300), the Shanghai Rising-Star Program (Grant No. 24QA2702100), and the Science and Technology Commission of Shanghai Municipality (Grant No. 22DZ2229004).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slc superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence* **34**(11), 2274–2282 (2012)
2. Ali, S., Jha, D., Ghatwary, N., Realdon, S., Cannizzaro, R., Salem, O.E., Lamarque, D., Daul, C., Riegler, M.A., Anonsen, K.V., et al.: A multi-centre polyp detection and segmentation dataset for generalisability assessment. *Scientific Data* **10**(1), 75 (2023)
3. Ayzenberg, L., Giryes, R., Greenspan, H.: Protosam for automated one shot medical image segmentation using foundational models. *Scientific Reports* **15**(1), 41482 (2025)
4. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilarinho, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics* **43**, 99–111 (2015)
5. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pragnet: Parallel reverse attention network for polyp segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. pp. 263–273 (2020)
6. Huang, C.H., Wu, H.Y., Lin, Y.L.: Hardnet-mseg: A simple encoder-decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps. *arXiv preprint arXiv:2101.07172* (2021)

7. Jain, S., Atale, R., Gupta, A., Mishra, U., Seal, A., Ojha, A., Jaworek-Korjakowska, J., Krejcar, O.: Coinnet: A convolution-involution network with a novel statistical attention for automatic polyp segmentation. *IEEE Transactions on Medical Imaging* **42**(12), 3987–4000 (2023)
8. Jha, D.: Kvasir-seg: A segmented polyp dataset. in multimedia modeling: 26th international conference, mmm 2020, daejeon, south korea, january 5–8 (2020)
9. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 4015–4026 (2023)
10. Liu, Y., Zhu, M., Li, H., Chen, H., Wang, X., Shen, C.: Matcher: Segment anything with one shot using all-purpose feature matching. *arXiv preprint arXiv:2305.13310* (2023)
11. Mao, X., Xing, X., Meng, F., Liu, J., Bai, F., Nie, Q., Meng, M.: One polyp identifies all: One-shot polyp segmentation with sam via cascaded priors and iterative prompt evolution. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 24182–24191 (2025)
12. Meng, L., Lan, S., Li, H., Alvarez, J.M., Wu, Z., Jiang, Y.G.: Segic: Unleashing the emergent correspondence for in-context segmentation. In: *European Conference on Computer Vision*. pp. 203–220 (2024)
13. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023)
14. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: *International conference on machine learning*. pp. 8748–8763 (2021)
15. Rahman, M.M., Marculescu, R.: Medical image segmentation via cascaded attention decoding. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 6222–6231 (2023)
16. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Koutra, C., Whitehead, S., Wang, X., Kirillov, A., Krahenbuhl, P., Feichtenhofer, C.: Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024)
17. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **71**(3), 209–249 (2021)
18. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. *IEEE transactions on medical imaging* **35**(2), 630–644 (2015)
19. Xu, Y., Tang, J., Men, A., Chen, Q.: Eviprompt: A training-free evidential prompt generation method for adapting segment anything model in medical images. *IEEE Transactions on Image Processing* **33**, 6204–6215 (2024)
20. Zauber, A.G., Winawer, S.J., O'Brien, M.J., Lansdorp-Vogelaar, I., van Ballegoijen, M., Hankey, B.F., Zauber, S.D., Burt, R.W., Bond, J.H., Lowery, M., et al.: Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *New England Journal of Medicine* **366**(8), 687–696 (2012)
21. Zhang, R., Jiang, Z., Guo, Z., Yan, S., Pan, J., Ma, X., Dong, H., Gao, P., Li, H.: Personalize segment anything model with one shot. *arXiv preprint arXiv:2305.03048* (2023)